

CS 134  
*Problem Set 2*

Xiner Zhou

February 8, 2017

**1. The square lattice. (20 points)**

- a. (6 points)** The pairs of nodes in a square lattice that are farthest apart (i.e. largest lattice distance) must be those that are located on the diagonals, say A and B, and they have  $2(r-1) = 2(\sqrt{n}-1)$  lattice distance apart, given that there exist short-cuts between every pair of nodes that are two lattice distance apart, the shortest path between A and B has length  $\frac{2(r-1)}{2} = \frac{2(\sqrt{n}-1)}{2} = \sqrt{n}-1$ . Therefore, the diameter of a 2-square lattice on  $n$  nodes is  $\sqrt{n}-1$ .
- b. (8 points)** Consider a node  $u = (i, j)$  that is at lattice distance at least 2 from any nodes on the edge of the lattice, the number of nodes that are  $d$  lattice distance away from  $u$  is  $4d$ . Therefore, the total number of neighbors  $u$  has is:

$$|N(u)| = \sum_{d=1}^2 4d = 4 + 8 = 12$$

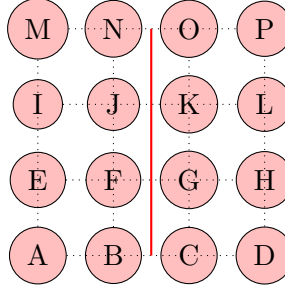
Now, let's calculate the number of edges among neighbors of  $u$ , i.e., how many neighbors of  $u$  are themselves neighbors. We observe that, the nodes that are 1 hop away from  $u$ , i.e.,  $(i-1, j), (i, j-1), (i+1, j), (i, j+1)$  each has 6 neighbors in  $N(u)$ ; among the nodes that are 2 hops away from  $u$ ,  $(i-2, j), (i+2, j), (i, j-2), (i, j+2)$ , each has 3 neighbors in  $N(u)$ ; among the nodes that are 2 hops away from  $u$ ,  $(i+1, j+1), (i-1, j-1), (i-1, j+1), (i+1, j-1)$  each has 6 neighbors in  $N(u)$ . Therefore, the total number of edges in  $N(u)$  is:

$$\frac{4(6 + 3 + 6)}{2} = 30$$

The clustering coefficient of a node  $u$  is:

$$c(u) = \frac{30}{\frac{12!}{10!2!}} = \frac{5}{11}$$

- c. (6 points)**



For even  $r$ , we could cut the square lattice as the above example shows, which cuts the graph into equal halves. For this kind of cut, we find a set of nodes  $S$  of size equal to  $\frac{n}{2}$  has  $r = \sqrt{n}$  edges leaving it, so by definition, the expansion parameter of the graph is smaller or equal to  $\frac{r}{\frac{n}{2}} = \frac{\sqrt{n}}{\frac{n}{2}} = \frac{2}{\sqrt{n}} = \frac{2}{r}$ .

**2. A smaller world? (Easley and Kleinberg, 20.8 Q1) (20 points)** We can use a  $k$ -square lattice on  $n$  nodes, where  $n$  is the full population of the world, and the lattice distance us a "closeness" or "friendship" measures that basically quantifies how close each pair of people are. For every node in the lattice, we create an edge to their closest 10 nodes in the lattice distance.

Notice that in a 2-square lattice, each node already has 12 neighbors, and in a 1-square lattice, each node has only 4 neighbors, therefore, the graph we are generating for the "close-friend" is less connected than a 2-square lattice, it should be somewhere between 1- and 2-square lattice.

From **1 a.**, we know that for  $k = 2$ , the diameter of the 2-square lattice is  $\sqrt{n} - 1$ . And we can easily observe that for 1-square lattice, the diameter is  $2(\sqrt{n} - 1)$ . Therefore, the "close-friend" network should have a diameter proportional to  $\sqrt{n}$ , or  $\Theta(\sqrt{n})$ . Since  $n$  is the total population of the world,  $\sqrt{n}$  is far much greater than 6. It's equivalent to say that there exists at least a pair of people in the world, that the path connecting requires at least  $\Theta(\sqrt{n})$  steps.

In conclusion, it is impossible that for each pair of people in the world, there is a path of at most six edges connecting this pair of people.

**3. Keep your friends close, and your acquaintances closer. (Easley and Kleinberg, 20.8 Q2) (20 points)**

**a. (8 points)** I expect the close-friend network to have the largest clustering coefficient.

To think about clustering coefficients, we are essentially talking about the principle of triadic closure: If two people in a social network have a friend in common, then there is an increased likelihood that they will become or already are friends themselves. The supporting arguments are based on opportunity, trust, and incentives. And furthermore, the likelihood of triadic closure are more powerful when the edges are strong ties (in this case, top 10 closest friends) than when they are weak ties (equivalent to 21-30 friends).

**b. (8 points)** I expect to see the "long-range" edges in distant-friend networks.

In the Watts-Strongatz model and the Kleinberg model, the lattice distance represent some kind of notion of similarity that guides the formation of the links. It could be geographic proximity or social proximity. By definition, for any people, we denote it as  $u$  in the graph, the 10 closest friends are more likely (or certainly) have small distance away from  $u$ , while the 21-30 friends are more likely to be not so much alike  $u$ , therefore, are more likely to have "long-range" edges away from  $u$ .

c. (4 points) I expect  $D$  to be larger.

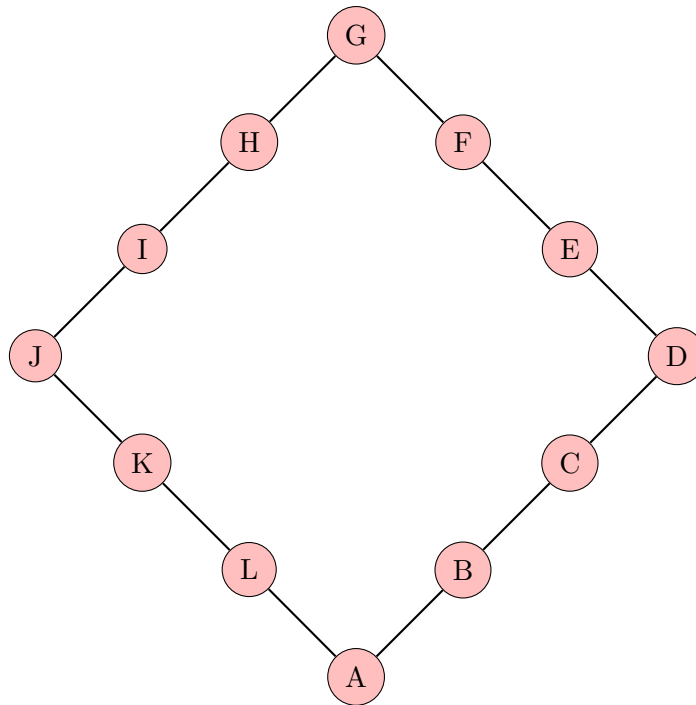
As I indicated in 3 a., close-friend network (strong ties) tend to have larger clustering coefficient than the distant-friend network (weak ties), that is to say, close-friend network has more triangles – sets of three people who mutually know each other. As a result, in each branching step, the number of nodes a person can reach in a close-friend network is less than that in a distant-friend network, because many of these edges go from one friend to another, not to the rest of the world, in a close-friend network. Therefore, the average number of people that a person can reach in six steps in the distant-friend network should be expected to be larger.

#### 4. Regular graphs. (20 points)

a. (6 points)

A ring lattice with  $n=12$  nodes and  $k=1$ . Justifications:

- Each pair of nodes has a path connecting them, and each node has 2 edges, so it is a connected 2-regular graph.
- The longest distance is between the nodes on the opposite side of a diameter, which is  $\frac{n}{2}$ . Therefore, the diameter is  $\frac{n}{2}$ .
- There is no triadic closure, so the clustering coefficient of the graph is 0.



b. (14 points)

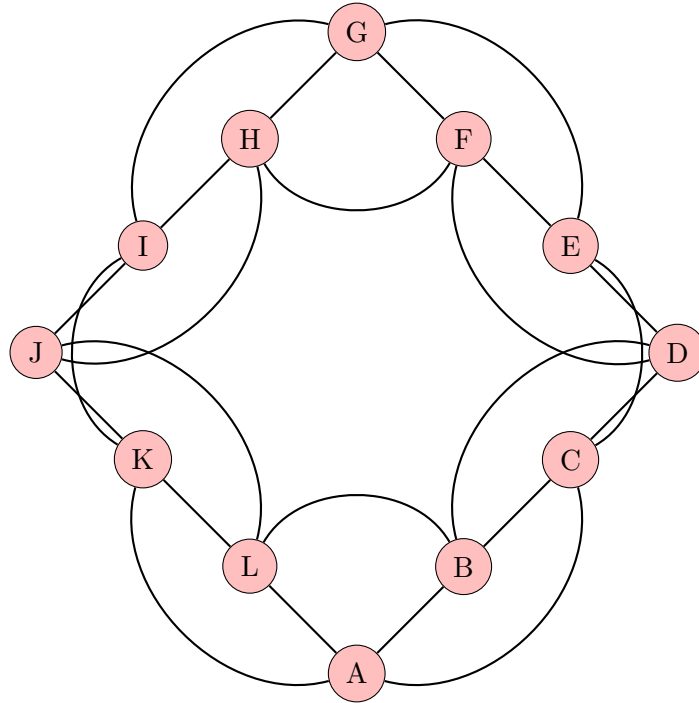
A ring lattice with  $n=12$  nodes and  $k=2$ . Justifications:

- Each pair of nodes has a path connecting them, and each node has 4 edges, so it is a connected 4-regular graph.

- The longest distance is between the nodes on the opposite side of a diameter, which is  $\frac{n}{2}$ . And since each pair of nodes that are 2 hops away are connected, the distance between them is  $\frac{n}{4}$ . Therefore, the diameter is  $\frac{n}{4}$ .
- For a node in the graph  $u$ , it has 4 neighbors, and only 3 pairs of its neighbors are themselves connected, therefore, the clustering coefficient for  $u$  is:

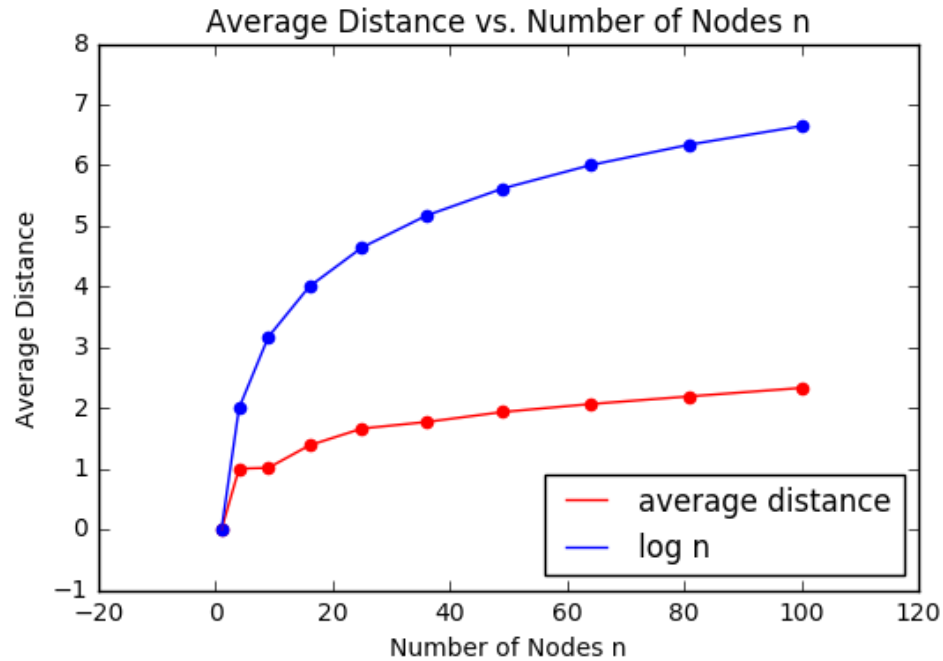
$$C(u) = \frac{3}{\frac{4!}{2!(4-2)!}} = \frac{1}{2}$$

The clustering coefficient is the same for all nodes in the graph, so the clustering coefficient of the graph is also  $\frac{1}{2}$ .

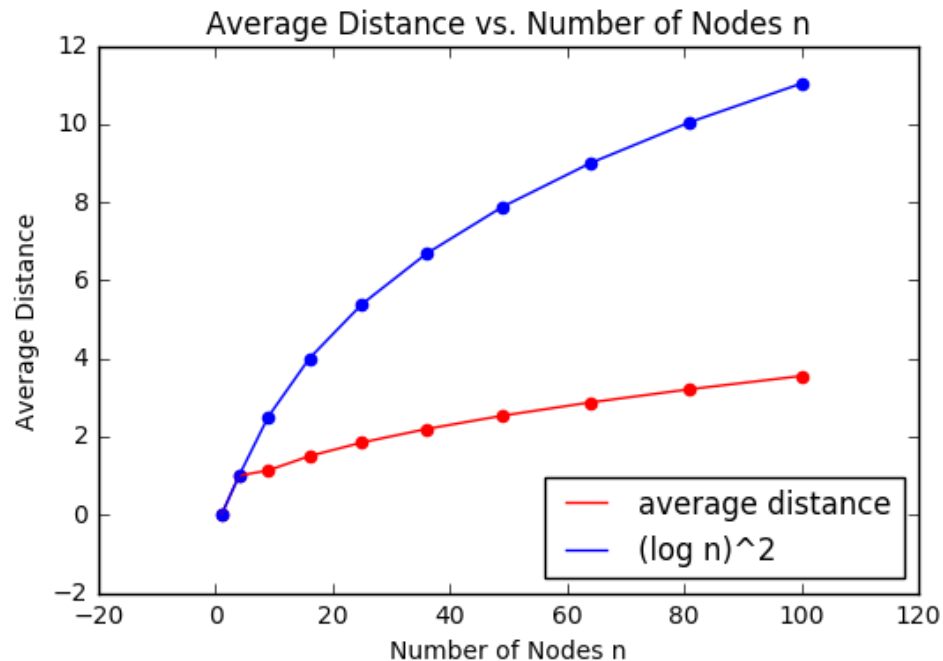


### 5. Coding: short distances in networks. (20 points)

- (7 points) For  $r = 1, 2, \dots, 10$  (or  $n = 1, 4, \dots, 100$ ), generate networks according to the Watts-Strogatz model with  $k = 2$  and  $l = 2$ . Plot the average distance between any two nodes as a function of  $n$  and plot the function  $\log n$ ;



- (6 points) For  $n = 1, \dots, 100$ , generate networks according to the Kleinberg model with  $k = 2$  and  $l = 2$ . Plot the average distance between any two nodes as a function of  $n$  against  $\log^2 n$ ;



- (7 points) Compute a **good estimate** (note: this does not need to be exact; clever solutions that provide good approximations of the shortest path are preferable to brute-force solutions) of the average shortest distances of these graphs.

Original network is too large for the Floyd-Warshall algorithm to work through efficiently. Therefore, I sample 100 nodes from the network, and make a sample subgraph, which should be a good estimate. The average distance (calculated by sampling) in the enron network and the epinions network are both infinity, that is, they are not connected. However, the livejournal network is too large to run on my computer. Please see my python notebook for details.